

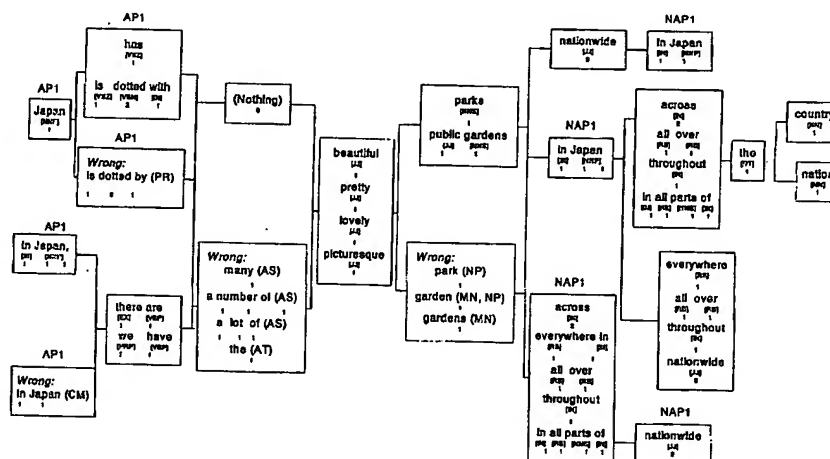


**PCT**



- 〔続葉有〕

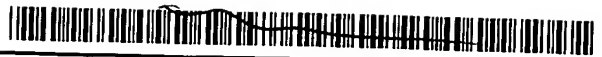
(54) 発明の名称: テンプレート-テンプレート構造に基づく対話式学習システム



(57) 要約: 本発明の学習システムにはテンプレート・オートマトンの概念が導入されており、「正しい」回答と「誤った」回答からなる多くの「多様な学習者の予想される例」を収集し、効率的なエラー診断エンジンとしてHCS(最重共通文字列)またはLCS(最長共通文字列)アルゴリズム等の代表的なNLP技術を言語学習システムに用い、テンプレート内に埋め込まれたこれらの例を学習者の回答の診断解析のために用いる。この診

〔続葉有〕

**WO 2004/084156 A1**



LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA,  
NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE,  
SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US,  
UZ, VC, VN, YU, ZA, ZM, ZW.

NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG,  
CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告書

- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC,

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

## 明 細 書

## テンプレートーテンプレート構造に基づく対話式学習システム

## 技術分野

本発明は、抽出ルール・ベースのテンプレートーテンプレート構造およびバギー・ルールの拡張性に富む機能を利用した新しい対話式学習システムに関する。

## 背景技術

言語指向インテリジェント学習システムに使用される時間の掛かるオーサリング・タスクの自動化と簡素化が切望されていることが発明の動機となっている。想定される模範回答数は合理的に制限するとしても、理想的な学習システムを開発しようとする場合、学習者が実際に犯す可能性のあるエラーは理論的には無限個に近い数にのぼる場合がしばしばあるからである。本発明者らが判断する限り、少なくとも予見できる将来においては、最先端の自然言語処理技術をもってしても、完全な自由形式でしかもエラーの多い文章の自動訂正を即座に提供出来るレベルには達していない。多くの有能な人間教師のようにシステムがこのような対処ができるには、いわゆる常識と呼ばれる世界の知識ベースをこのシステムに導入することができる場合にのみ可能であろう。

## 発明の開示

本発明の学習システム(Azalea)にはテンプレート・オートマトンの概念が導入されており、「正しい」回答と「誤った」回答からなる多くの「多様な学習者の予想される例」を収集する。言語学習システムで使われる効率的なエラー診断エンジンとして HCS(最重共通文字列)または LCS(最長共通文字列)アルゴリズムという代表的な NLP 技術が決定的な役割を担っており、テンプレート内に埋め込まれたこれらの例は学習者の回答の診断解析のために使用される。この診断は、テンプレート・データベースの膨大な数の候補パスの中から学習者の入力文に最も近似度の高いパスを選択することによって行われる。適格なモデル翻訳文と不適格な誤った文章から構成されるテンプレ

ート・コーパスを構築するオーサリング・タスクは、多くの時間を費やし、人手を浪費させる。

本発明の新しいシステムには、通常であれば時間の掛かるテンプレート生成のオーサリング・タスクを簡素化、すなわち低減することにおいてのみならず(徳田 尚之、陳亮、笹井 紘幸らによる特開 2002-49617 号等を参照のこと)、システム性能の向上においても有効である。導入されたテンプレート・アーキテクチャがシステムの簡素化と性能の向上をもたらす第 1 の理由は、このアーキテクチャが、単一のテンプレート・アーキテクチャの遷移ノードのいくつかに割り当てられた抽出ルールを適用することによって、単一のテンプレート・アーキテクチャに多くの異なるテンプレートを統合したり、また逆にいうとそこから多くの異なるテンプレートを同様に抽出することが可能になることであり、第 2 の理由は、導入されたバグ・ルールが学習者の誤った回答を自動的に識別・分類し、したがってそこからバグを生成する機能を有することである。NLP (自然言語処理) 技術がこのシステム開発で主要な役割を担う理由は、学習者の自由形式回答の構文構造を調べるためにはパーサが使用されておき、その意味論的構造は、学習者からの回答を準備されたテンプレート・データベースの意味論的に等価のパスとの文字列照合により調べることから明らかであろう。

新しい抽出ルール・ベースの、またバグ・ルール・ベースのテンプレート・アーキテクチャ構造は、文字入力および会話方式による学習システム、音声利用技術によるコール・センターまたは音声ポータル・システム、あるいはシステムと人間との間でより自然な人間とコンピュータの対話を実施するより強化されたヒューマン・コンピュータ・インターフェースに焦点を合わせたシステムならばどのシステムをも含めて多くのアプリケーションで重要な役割を担うことが期待される。

この発明では以下のような機能が与えられる。

1. ある制約条件の下で選択したテンプレート・アーキテクチャ遷移図のノードに抽出ルールを埋め込むと、単一テンプレート・アーキテクチャは様々な異なるタイプの既存のテンプレートを表現することができる。
2. 導入されたバグ・ルールの拡張性に富む機能を利用することによって、多くの

誤り表現および／または多くの誤った構文構造は、バグ分類学を記述するその自動機能によって、新しいテンプレートテンプレートを「よりファットな」テンプレートテンプレート形式に展開できるように整合性を保って自動的に記述することができる。語学教師は、テンプレートテンプレートを制作する際に誤った翻訳文の分類に費やす必要はなく、したがってエラー分類学の負担を低減し、学習者らの特定のエラーの特徴を評価しクラスタ化する。

3. HCS マッチング・アルゴリズムは、入力文をより簡素なテンプレートテンプレートに直接照合し、したがって実際にテンプレートテンプレートを展開せずにすべての抽出されたテンプレートのすべての予想されるパスから最適なパスを探索する際の照合処理の空間および時間計算量を低減するように開発することができる。

発明を実施するための最良の形態

テンプレートテンプレート構造

まず「テンプレートテンプレート」という用語を以下の様に定義する。テンプレートテンプレートは、ノードのいくつかは抽出ルールに関連付けられたシンボルでマーク付けされており、そのテンプレートテンプレートが多くのテンプレート、または一組の非連結型テンプレートが1つのテンプレートとみなされる場合はいわゆる大きいテンプレートに展開することができる場合の特別なテンプレートと定義される。このような一組の切断されたテンプレートによって、単一の L1 文章の様々な予想される翻訳文は、一群の翻訳された L2 文章から構成される大きな単一のテンプレートテンプレートを形成することが可能になる。展開テンプレートであるので、テンプレートテンプレート・スキームによってテンプレートテンプレートはいわば1つまたは複数のテンプレートを抽出することが可能になる。

通常、抽出ルールは常に一組のシンボル、例えば $\{s_1, s_2, \dots, s_n\}$ に関連付けられており、そのシンボルのそれぞれにはテンプレートの1つまたは複数のノードが割り当てられている。それらの関連付けられたシンボルには1つまたは複数の値が割り当てられており、その機能は、テンプレートテンプレートから抽出された1つ以上のテンプレートに現れるノードのスタイルを表現することである。本発明ではこれらのシンボルを「ラベ

ル・シンボル」と呼ぶ。単一ルールに関係付けられたシンボルは「関係シンボル」と呼ばれる。関係シンボルはある種の制約を有するべきである。一般的な制約としては、所与の  $s_i=1$  に際して、 $s_k$  はしばしば 2 に制約されるか、あるいは 1 以外のいくつかの正の整数に制約される必要がある。 $s_i$  の値が一組の他のシンボルに割り当てられた値に依存する場合、 $s_i$  の値の選択は他のシンボルの必須選択肢と呼ばれる。

語学教師が容易に理解できるように抽出ルールのいくつかの例を以下に示す。

#### タイプ A ルール AP(出現)– NAP (非出現)ルール

ノードによって  $AP_i$  でマーク付けされているノードと、 $NAP_i$  でマーク付けされているノードがあると想定する( $i$  は任意の整数で別のタイプの A ルールを示す)。タイプ A ルールの AP–NAP ルールは、展開された時に新たに展開されたテンプレートは  $AP_i$  でマーク付けされたノードかまたは  $NAP_i$  でマーク付けされたノードを含むことができるが、これらの両方を同時に含むことはできないという条件を課す。本発明においては、 $AP_i$  でマーク付けされたノードがテンプレートに現れないという場合を示すために  $AP_i=0$  を使用する。この時、 $NAP_i$  は、 $NAP_i$  でマーク付けされたノードがそのテンプレートに現れるであろうことを意味する 1 である必要がある。したがって本発明では  $NAP_i=1$  は  $AP_i=0$  の必須選択肢であると言えることができる。同様の理由から、 $NAP_i=0$  の場合、 $NAP_i=0$  が  $AP_i=1$  の必須選択肢となるように、 $AP_i$  は 1 の値を有する必要がある。

#### タイプ B ルール PPR (人称代名詞) – PPRP (人称代名詞所有格)ルール

タイプ A ルールと同様に、テンプレート–テンプレート・ルールは、一組のテンプレートに現れる  $PPR_i$  でマーク付けされたノードと  $PPRP_i$  でマーク付けされた他のノード( $i$  は任意の整数)は、それぞれに代名詞の自然言語文法が必要とする人称代名詞のフォームと代名詞の人称代名詞所有格のフォームでなければならないという条件を課す。 $PPRP_i$  (または  $PPR$ ) の場合、 $PPR_i$  (または  $PPRP_i$ ) に要求される値は代名詞の自然言語文法によって定義される必要がある。

#### タイプ C ルール AN (任意数)ルール

タイプ C ルールは、 $AN_i$  でマーク付けされたノードには任意のポジティブな実数を割り当てることができるという条件を課す。「I have 5 books on Zen」が真ならば、5 以外の数はどの数も誤りなので、このルール  $AN_i$  はエラー・ノード 5 に割り当てることができる。

#### タイプ D ルール CHO (択一)ルール

タイプ D ルールは、 $CHO_{i_1}$ ,  $CHO_{i_2}$ , ...,  $CHO_{i_k}$  でマーク付けされたテンプレート・テンプレートのすべてのノードの中で、テンプレート・テンプレートから抽出されたどのテンプレートにもただ一組だけのノードしか現れることができないという条件を課す。ここで異なる  $i$  は異なるタイプ D ルールを表している。したがって、 $CHO_{i_j}=0$  は  $CHO_{i_j}$  でマーク付けされたノードは現れないことを意味し、 $CHO_{i_j}=1$  はその指定されたノードが現在現れていることを意味している。明らかに、1 が 1 つの  $CHO_{i_k}$  に割り当てられている場合、0 は他のすべての  $CHO_{i_j}$  に割り当てられるべきである。

#### テンプレート・テンプレートを拡張するためのバギー・ルール

ここでいうバギー・ルールは、正確な構文規則からの予想される逸脱によって特徴付けられる一般的な構文上の誤り表現のプロダクション・ルールと定義される。具体的に、バギー・ルールの以下のフォームを想定する。

$$H_1 H_2 \dots H_N \rightarrow R_1 R_2 \dots R_M$$

ここで  $H_1 H_2 \dots H_N$  は任意のテンプレート・テンプレートの構文上正確なパスを追跡する一組のノードまたは正確な表現の基本的な構成要素またはセグメントを表す一組の文法的な品詞タグを表す。 $R_1 R_2 \dots R_M$  は、正確なフォームが  $H_1 H_2 \dots H_N$  である典型的誤り表現を表すノードのセットである。エラーがテンプレート・テンプレートの正確なパスからのずれによって識別されることが即座に理解される。一例を示す。

EX VBP  $\rightarrow$  EX VBZ

(ここで EX は「there is」(～がある)というような存在を表すものであり、VBP は 1 人称および 2 人称の現在形の動詞であり、VBZ は 3 人称単数現在形の動詞である)。この例は、構文上正確な表現である「there are 5 books」が、主語と動詞の属性一致を誤

って理解している生徒によって誤用されており、この例では誤り表現である「*there is 5 books*」が生じたことを意味している。

#### 図面の簡単な説明

図1は本発明のテンプレートーテンプレートの構造を示す図である。

図2は展開ルールにより展開されたテンプレートーテンプレート構造を示す図である。

図3は「Japan is dotted with beautiful parks nationwide」という意味の文章の例で展開されたテンプレート1を示す図である。

図4は「Japan is dotted with beautiful parks nationwide」という意味の文章の例で展開されたテンプレートを示す図である。

#### 【符号の説明】

エラー・メッセージ:

AS: 名詞の量に仮定が行われている、

AT: 冠詞は必要ない、

CM: コンマが必要である、

CT: 短縮形が不正確である、

MN: 意味が不正確である、

NP: 名詞は複数形でなければならない、

VS: 主語が単数形なので動詞は単数形でなければならない、

PR: 前置詞が不正確である、

PP: 句は複数形である必要がある、

一般的な品詞タグ:

DT: 限定詞

IN: 前置詞／従属接続詞

NN: 名詞、単数または集合

NNP: 固有名詞、単数

VBN: 動詞、過去分詞

VBZ: 動詞、3人称単数、現在形

EX: 存在を表す

JJ: 形容詞

NNS: 名詞、複数

RB: 副詞

VBP: 動詞、非3人称単数、現在形



以下本発明を図面に示す具体例によって説明する。

図1はオリジナルなテンプレートーテンプレートの構成を示す図であり、図2は上記の展開ルールにより展開されたテンプレートーテンプレートを示す図、図3はテンプレート1を示す図であり、図4はテンプレート2を示す図である。

テンプレートーテンプレート、抽出ルールによるテンプレート拡張、およびバギー・ルールの例

本発明の実施例においては、まず「Japan is dotted with beautiful gardens nationwide.」という意味の日本語の文章の英語への翻訳文に関するテンプレートーテンプレートを構築する。図1等示される番号は、文中における各単語の相対的重要性を強調する各単語の重みを示している。テンプレートの単語のデフォルトの重みは1にセットされており、これらはこの分野の専門家によって判断されるその単語の重要性に関連付けて割り当てられる必要がある。徳田 尚之、陳 亮、笹井 紘幸による特開2002-49617号を詳細な説明のために参照する。「[」と「]」の間のシンボルは品詞タグである。図中左端に示すノードは開始ノードである。

前記バギー・ルールを単純に適用することによって、それを図2のテンプレートーテンプレートに展開することができる。

これは、前記のバギー・ルールがバグの分類法を生成することができ、それによってそれらの誤り表現をテンプレートーテンプレートに構築することを自動的に可能にするので、語学教師がテンプレートーテンプレートを構築する際に多くの一般的なエラーを分類することの詳細に配慮する必要はないことを示している。

次に、前記タイプAルールを適用すると、 $AP_i$ でマーク付けされたノードがそのテンプレートに現れることを可能にし、したがって図2の $NAP_i$ でマーク付けされたノードを削除することによって図2のテンプレートーテンプレートから図3のテンプレート1を抽出し、ならびに図2の $AP_i$ でマーク付けされたノードを削除し、したがって $NAP_i$ でマーク

付けされたノードがそのテンプレートに現れるようにすることによって図4のテンプレート2を抽出することができることが容易に分かる。

語学教師がより簡素なラベル・シンボルに関してテンプレートの1つの大きな組み合わせを完成させてテンプレートーテンプレートを構築することができることが理解される。

テンプレートーテンプレートおよび入力文に対する最重共通文字列マッチング・アルゴリズム

前記の詳細な説明から明らかなように、単一テンプレートーテンプレートから多くのテンプレートを抽出することができる。テンプレートーテンプレートが遷移図上のあるノードに関連付けられるべきラベル・シンボル  $s_1, s_2, \dots, s_n$  を有すると仮定すると、テンプレートーテンプレートから抽出された異なるテンプレートはそれらのシンボルをノードに割り当てることによって獲得することができる。本発明ではテンプレートから抽出された各テンプレートを  $n$  タプル  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  で示すことができる。ここで、 $p_i$  はシンボル  $s_i$  に対する適切な割り当てである。上記のセクションで既に議論した通り、 $p_i$  は使用される抽出ルールに従って数字または単語のどちらかであってもよい。

2つの文章の最重共通文字列は  $a_1$  の次に  $a_2$ 、その次に...、その次に  $a_m$  という順番で両方の文章に現れる順番付けされた単語の最重共通文字列  $a_1, a_2, \dots, a_m$  であると定義される。共通文字列の定義は、A. V. Aho および J. D. Ullman 著 (Computer Science Press, 1992, pp. 321-327) の教科書『*Foundations of Computer Science* [情報科学の基礎]』に記載されている。

テンプレート内の各単語または各語句には重みが割り当てられているので、テンプレート内のパスと入力文の最重共通文字列は、その合計の重みが最大であるすべての予想可能な共通文字列内の最重共通文字列と定義される。

次に、テンプレートのすべての予想可能な有効パスから入力文の単語および／または語句の最重共通文字列を検索する。

テンプレートおよび入力文の最重共通文字列は、それぞれがテンプレートと入力文1つのパスから獲得される最重共通文字列中で最も重い合計重量を有する単語の最

重共通文字列と定義される。

言語翻訳学習システムに関するアプリケーションでテンプレートーテンプレートが獲得されると、次のステップは、入力文をすべての予想可能なテンプレートのそれぞれと照合し、次いで最も近いパスを選択することである。文章に対するテンプレートの DP (ダイナミックプログラミング) ベースの照合手順の詳細な説明は、徳田 尚之、陳 亮、笹井 紘幸らによる特開 2002-49617 号等に記載されている。

本発明の方法では、テンプレートーテンプレートからすべてのテンプレートを物理的に抽出せずに、抽出ルールによって(しかしバギー・ルールにはよらずに)直接的に抽出することができるテンプレートのすべての有効なパスの中から最も近いパスを探索する。そのような照合を行う前にテンプレートーテンプレートにはいかなるバギー・ルールをも含めないように、まずバギー・ルールが埋め込まれたテンプレートーテンプレートを展開する必要がある。これはすでに述べた図 2 のステップに従って実行することができる。

アルゴリズムで必要とされる第 1 のステップは、適用可能な空ノードごとに  $\square$  というラベルの弧に重み 0 を加算してテンプレートの各ノードをグラフの 1 つまたは複数の弧として単純に表現することによって、テンプレートーテンプレートを非循環的な重み付けされた有限有向グラフ(有向グラフ)の 2 重数値に変換することである。有向グラフはテンプレートーテンプレートから変換されるので、その関数とそのシンボルに割り当てられた値に決定的に依存するラベル・シンボルに関連付けられた多くの弧を含む。したがってそのような 1 つの有向グラフがあると仮定すると、異なる一組のラベル・シンボルが弧に割り当てられている場合は完全に異なるテンプレートを抽出することができる。すなわち、そのような有向グラフがあると仮定すると、テンプレートーテンプレートから抽出することができるテンプレートにそれぞれが対応する多くの有向グラフを獲得することができる。テンプレートーテンプレートから抽出された有向グラフを、以後、テンプレート有向グラフと呼ぶ。

本発明者らは、次にすべての有向グラフのパスと入力文の共通文字列から最重共

通文字列を探索する手順を以下で定義する。

有向グラフの任意の特定のノード  $N$  で終了するパスと入力文の最重共通文字列は、有向グラフ  $N$  で終了する 1 つのパスと入力文からそれぞれが獲得されるすべての最重共通文字列中最も重い合計重量を有する単語のシーケンスと定義される。

さらに、テンプレート有向グラフから抽出され、しかしノード  $N_i$  で終了するすべての有向グラフのパスを  $N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  で表現する。ここでシンボル  $s_i$  は値  $p_i$

( $i=1, 2, \dots, n$ ) に割り当てられる。n タプル  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  をノード  $N_i$  のラベルと呼ぶ。ここで  $s_1$  を  $p_1$  に、 $s_2$  を  $p_2$  に、 $\dots$ 、 $s_n$  を  $p_n$  にセットするときにルールの矛盾がないものと仮定すべきである。このようなラベル  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  を矛盾なしラベルと呼ぶ。

$N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  でラベルされたノードと入力文の最重共通文字列は、矛盾なしラベル  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  でマーク付けされたノードを有する有向グラフ・テンプレートから抽出された 1 つの有向グラフの最重共通文字列としてそれぞれが獲得されるすべての最重共通文字列中で最も重い合計の重みを有する単語の最重共通文字列と定義される。

ノードによっては、1 つの有向グラフの AP2 でラベルされたノードと NAP2 でラベルされたノードのように有向グラフ・テンプレートから抽出された 1 つの有向グラフで同時に現れない場合があることに留意されたい。その結果、 $N_i(\dots, AP2, 1, \dots, NAP2, 1\dots)$  のようにルール違反のラベルは、有向グラフ・テンプレートのノードと入力文の共通文字列のどの計算計画にも入ることを許可されるべきではない。以下のアルゴリズムはテンプレート・テンプレートと入力文の最重共通文字列を計算するための手順を記述する。以下の計算では、ラベル・シンボルの非常に特別の値として「 $\lambda$ 」が使用され、それによってその値は計算の特定の段階に達するまで未指定のままである。

1. テンプレート・テンプレートを、その有向エッジ(遷移)がテンプレートの対応する単語によってラベルされるテンプレート有向グラフに変換する。
2. 有向グラフのすべてのノードを、ノード  $N_i$  および  $N_j$  の対ごとに  $j > i$  の場合に  $N_j$  から  $N_i$  への遷移がないように  $N_1, N_2, \dots, N_t$  のノードにトポロジカルに分類する。

3. 空のノード  $N_0$  を有向グラフに追加し、 $N_0$  からの弧をテンプレートーテンプレートのすべての開始ノードに加える。
4.  $CM(N_0, M_0) = 0$  にセットする。
5.  $i=0$  から  $t$  の場合、以下のステップを実行する。
6. 1 つのシンボルに関連付けられた  $N_i$  に対して少なくとも 1 つの弧があり、ノード  $N_i$  以降に  $s$  の関係ラベルがまったく存在しない場合、すべての  $i$  に関して  $j=0$  から  $m$  の場合は以下を実行する。

$s$  関係ラベルが  $\{s_1, s_2, \dots, s_n\}$  に現れておらず少なくとも 1 つの  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, sx_1, px_1, sx_2, px_2, \dots, sx_h, px_h\}$  が存在する場合、任意のラベル  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  に関してすべての  $CM(N_i[\dots], M_j)$  をチェックする。ここで  $CM(N_i[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, sx_1, px_1, sx_2, px_2, \dots, sx_h, px_h\}, M_j])$  が定義されるように、 $sx_1, sx_2, \dots, sx_h$  は  $s$  関係ラベルである。

$CM(N_i[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j])$  を最大の  $CM(N_i[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, sx_1, px_1, sx_2, px_2, \dots, sx_h, px_h\}, M_j])$  と定義し、すでに定義されているすべての  $CM(N_i[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, sx_1, px_1, sx_2, px_2, \dots, sx_h, px_h\}, M_j])$  を定義解除する。

7.  $j=0$  から  $m$  の場合、以下のステップを実行する。
8.  $N_i$  からの弧があるノード  $N_k$  のそれぞれに対しては以下を実行する。
  - (1) 弧  $N_i N_k$  がラベルを有しない場合、すでに定義されている  $CM(N_i[\dots], M_j)$ 、 $CM(N_i[\dots], M_{j+1})$ 、 $CM(N_k[\dots], M_j)$ 、 $CM(N_k[\dots], M_{j+1})$  をすべてチェックし、 $CM(N_i[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j])$ 、 $CM(N_i[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$ 、 $CM(N_k[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j])$ 、 $CM(N_k[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$  の 1 つが既に定義されている場合は  $CM(N_k[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$  を以下のデータの最大と定義する。
    - 既に定義されている場合は  $CM(N_i[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j])$  である。
    - $CM(N_i[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j])$  が既に定義されており、弧  $N_i N_k$  が  $W_k$  と一致する場合は  $M(N_i[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j]) + W(N_i N_k)$  である。
    - 既に定義されている場合は  $CM(N_i[\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$  である。

- 既に定義されている場合は  $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$  である。
  - 既に定義されている場合は  $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$  である。
- (2) 弧  $N_i N_k$  がシンボル  $s$  に関連付けられている場合、既に定義されている  $CM(N_i\{\dots\}, M_j)$ ,  $CM(N_i\{\dots\}, M_{j+1})$ ,  $CM(N_k\{\dots\}, M_j)$ ,  $CM(N_k\{\dots\}, M_{j+1})$  をチェックする。
- (i) ノード・ラベル  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}$  が矛盾なしラベルであり、以下の少なくとも 1 つが既に定義されている場合:
- $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ ,  $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$
  - $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ ,  $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$
  - $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}, M_j)$ ,  $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}, M_{j+1})$
  - $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}, M_j)$ ,  $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}, M_{j+1})$
  - $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}, M_{j+1})$  を上記で定義されたデータの最大のデータと定義する。
- (ii) ラベル  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  があると仮定して、ラベル  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}$  が矛盾なしラベルであり、以下の少なくとも 1 つが真である場合:
- $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$  がすでに定義されており、 $s$  から  $p$  までのセッティングが  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  の必須選択肢であるか、 $p$  が  $s$  に割り当てられた後で  $N_i N_k$  が  $M_{j+1}$  と一致する場合
  - $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$  がすでに定義されており、 $s$  から  $p$  までのセッティングが  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  の必須選択肢の場合
  - $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$  がすでに定義されており、 $s$  から  $p$  までのセッティングが  $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$  の必須選択肢である場合
  - $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_{j+1})$  が既に定義されている場合
  - $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_j)$  が既に定義されている場合
  - $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_{j+1})$  が既に定義されている場合
  - $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_j)$  が既に定義されている場合
  - $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_{j+1})$  が既に定義されている場合

- ・  $CM(N_k\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_{j+1})$  を上記で定義されたデータと以下のデータの最大データと定義する。
  - ・  $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$  が定義されており、 $p$  が  $s$  にセットされた後で  $N_i N_k$  が  $M_{j+1}$  に一致する場合、 $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j) + W(M_{j+1})$  である。
  - ・  $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_j)$  が定義されており、 $p$  が  $s$  に割り当てられた後で  $N_i N_k$  が  $M_{j+1}$  と一致する場合、
  - ・  $CM(N_i\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_j) + W(M_{j+1})$  である。
- また、 $s$  として割り当てられた  $p$  に従い、これらの  $s_i$  を  $\lambda$  から値の必須選択肢まで変更する。

$N_x$  を最終的な頂点であると既に定義しているすべての  $CM(N_x, M_m)$  の中で最大の  $CM(N_x, M_m)$  はテンプレートーテンプレートおよびパスの最重共通文字列の重みになる。

上記のアルゴリズムでは、いくつかの候補からいつ  $CM(N(.), M.)$  を選択したとしても選択したその 1 つに対して一種のバック・リンクをセットすることに留意されたい。テンプレートーテンプレートとパスの最長共通文字列の重みを探索したように、このバック・リンクを追跡することによって、最重共通文字列を有する抽出されたテンプレートのパスを入力文と共に即座に獲得することができることに留意されたい。

#### 産業上の利用可能性

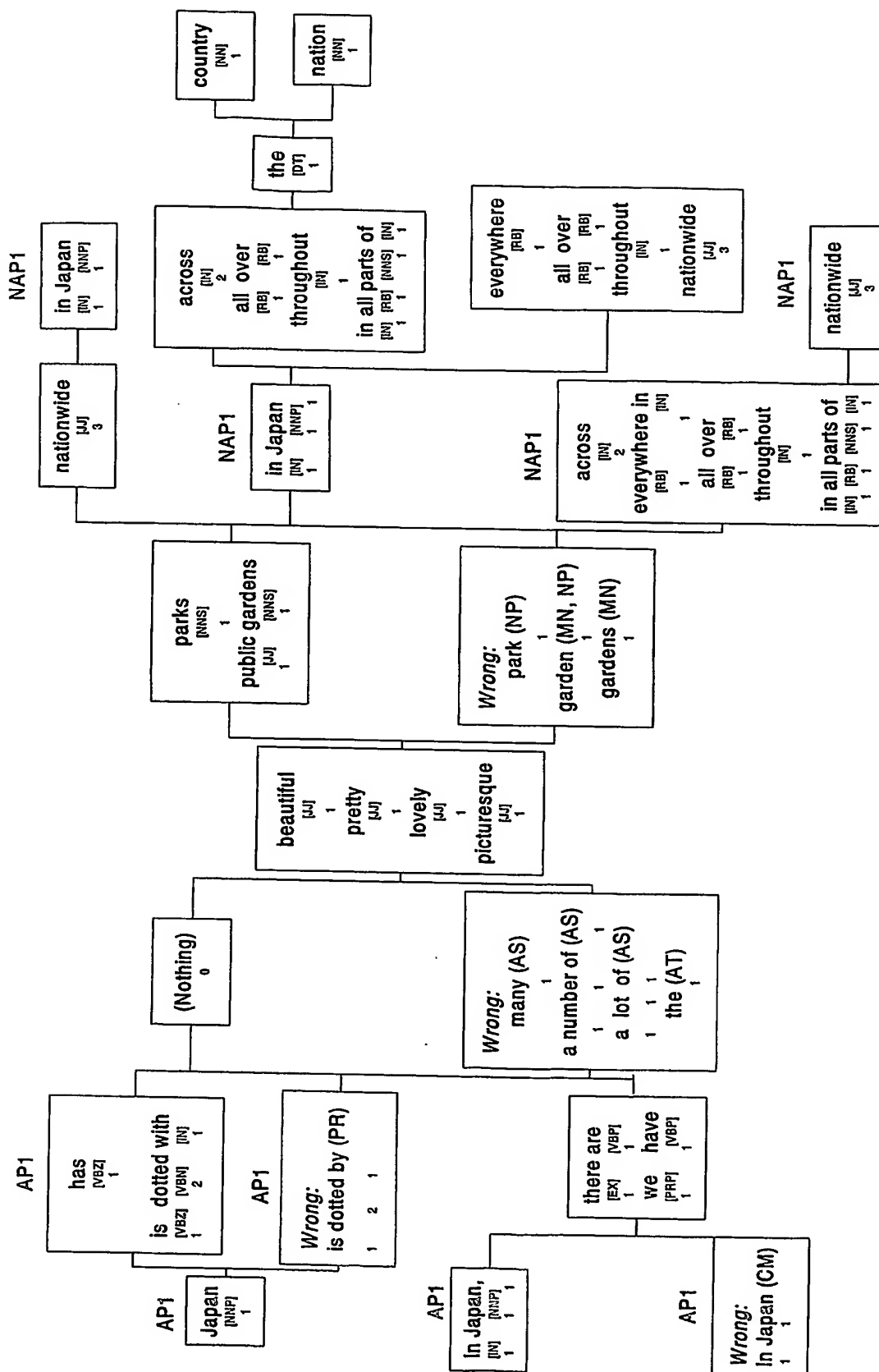
本発明を明細書中では代表的な応用例である自然言語翻訳の技術分野に関して記載したが、本発明の用途は自然言語学習システムに限定されるものではなく、本発明は音声利用技術、プログラミング言語学習システム、または例えば人間とコンピュータの対話を可能にするより自然な拡張型インターフェースを必要とするシステムであればどのようなシステムのためにも利用することができる。

## 請求の範囲

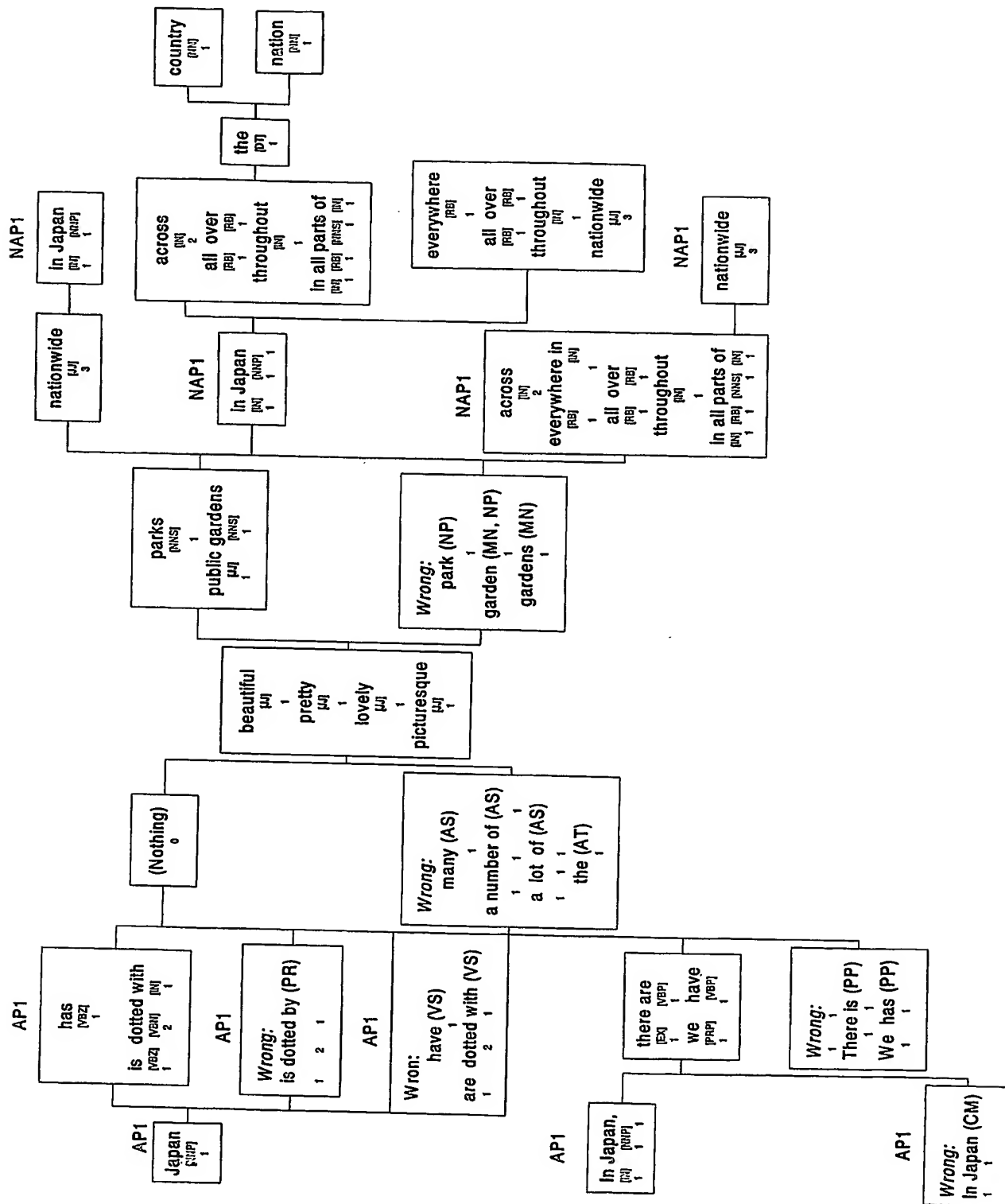
1. 抽出ルール・ベースのテンプレートーテンプレート構造およびバギー・ルールの拡張性を用いた対話式学習システムにおいて、テンプレートーテンプレート構造内のノードのいくつかは、多くのテンプレートまたはいわゆる大きいテンプレートに抽出することができる抽出ルールに関連付けられたシンボルでマーク付けされるテンプレートーテンプレート構造を用いた対話式学習システム。
2. 展開テンプレートのノードのいくつかに抽出ルールに関連付けられた特別なシンボルが割り当てられており、これによってテンプレートの 1 つまたは複数を抽出することができるようにされたテンプレートーテンプレート・スキームを用いる請求項 1 記載の対話式学習システム。
3. 単一テンプレートーテンプレートから複数のテンプレートを抽出するためにテンプレートーテンプレート内のノードをマーク付けするために使用される抽出ルール基く請求項 1 記載の対話式学習システム。
4. スリムなテンプレートーテンプレートをより大きなテンプレートーテンプレートまたはよりファットなテンプレートーテンプレートに展開するために用いられるバギー・ルール基く請求項 1 記載の対話式学習システム。
5. 前記抽出ルールが常に一組のシンボル、すなわち  $\{s_1, s_2, \dots, s_n\}$  に関連付けられ、このシンボルのそれぞれはテンプレートの 1 つまたは複数のノードに割り当てられ、これらの関連付けられたシンボルには 1 つまたは複数の値が割り当てられてなる抽出ルール基く請求項 1 記載の対話式学習システム。
6. 入力文を抽出ルールが埋め込まれたテンプレートーテンプレートと照合するための HCS マッチング・アルゴリズムを用いる請求項 1 記載の対話式学習システム。



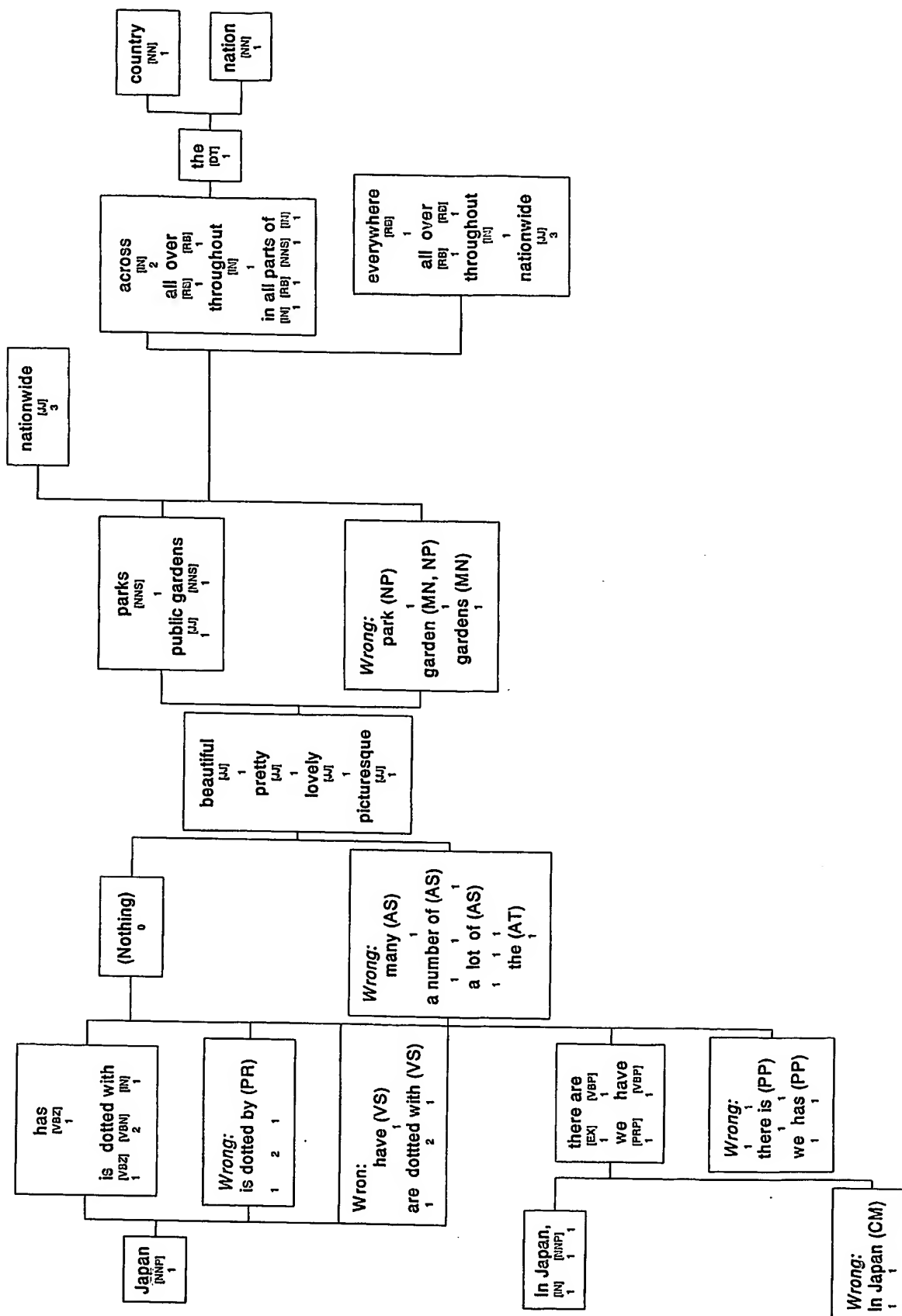
【図 1】



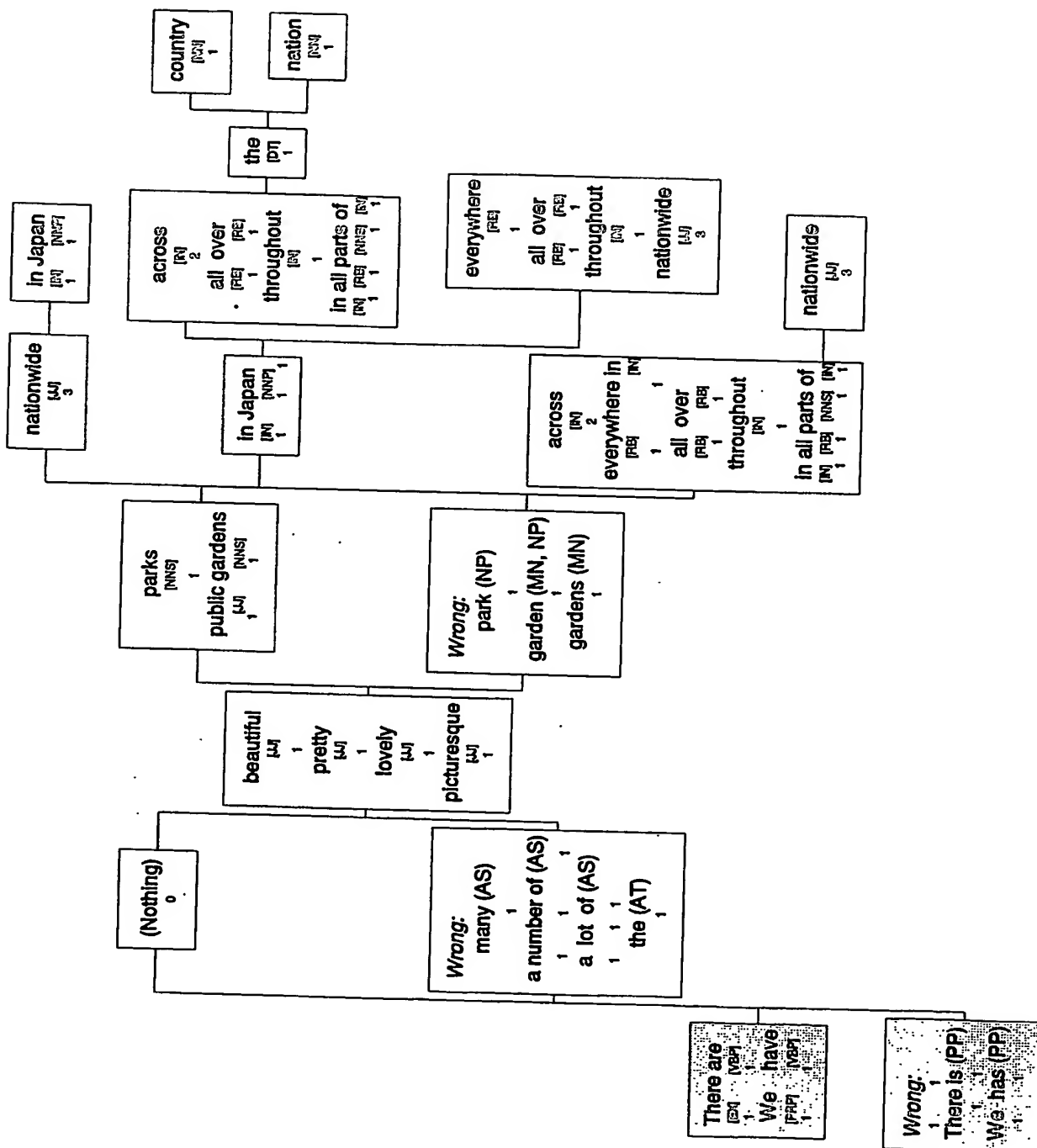
【図 2】



【図 3】



【☒ 4】



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/003838

## A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl<sup>7</sup> G09B19/06

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl<sup>7</sup> G09B19/06

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-2004

Kokai Jitsuyo Shinan Koho 1971-2004 Jitsuyo Shinan Toroku Koho 1996-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Naoyuki TOKUDA et al., "Template Automaton ni yoru On-line Chiteki Eisakubun Kyoiku Shien System", The Institute of Electronics, Information and Communication Engineers Ronbunshi, J84-D-I, Vol.7, 01 July, 2001 (01.07.01), Vol.J-84-D-I, No.7, pages 1089 to 1101, ISSN 0915-1915	1-6
Y	US 2001/044098 A1 (Benny G. JOHNSON et al.), 22 November, 2001 (22.11.01), Full text; Figs. 1 to 3 (Family: none)	1-6
E,A	JP 2003-150584 A (NEC Corp.), 23 May, 2003 (23.05.03), Full text; Figs. 1 to 23 (Family: none)	1-6

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

## \* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;" document member of the same patent family

Date of the actual completion of the international search  
21 June, 2004 (21.06.04)Date of mailing of the international search report  
06 July, 2004 (06.07.04)Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G09B19/06

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G09B19/06

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1922-1996年  
 日本国公開実用新案公報 1971-2004年  
 日本国登録実用新案公報 1994-2004年  
 日本国実用新案登録公報 1996-2004年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	徳田尚之他, テンプレートオートマトンによるオンライン知的英作文教育支援システム, 電子情報通信学会論文誌 J84-D-I 第7号, 01.07.2001, 第J84-D-I巻 第7号, p. 1089-1101, ISSN 0915-1915	1-6
Y	US 2001/044098 A1 (Benny G. JOHNSON et al.) 22.11.2001, 全文, 図1-3 (ファミリーなし)	1-6
E, A	JP 2003-150584 A (日本電気株式会社) 23.05.2003, 全文, 図1-23 (ファミリーなし)	1-6

☐ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの

「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」口頭による開示、使用、展示等に言及する文献

「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&amp;」同一パテントファミリー文献

国際調査を完了した日

21.06.2004

国際調査報告の発送日

06.07.2004

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

松川 直樹

2T

8804

電話番号 03-3581-1101 内線 3264